

**LAPORAN PENELITIAN**  
**SKEMA PENELITIAN**  
**KEILMUAN - KOM - Dosen Pemula**

**AREA PENELITIAN :**  
**Sains dan Teknologi era industri 4.0 dan masyarakat 5.0**

**INOVASI SOFTWARE EVALUATION PADA APLIKASI MOBILE MELALUI  
KOMBINASI USABILITY TESTING DAN SENTIMENT ANALYSIS**



Oleh :

**Ketua Peneliti:**

Hasan Basri, S.Kom., M.Kom.  
NIP. 199205172024061003

**Anggota :**

Irpan Kusyadi, M.Kom.  
NIP. 199010112024061002

**PROGRAM STUDI SISTEM INFORMASI**  
**FAKULTAS SAINS DAN TEKNOLOGI**  
**UNIVERSITAS TERBUKA**  
**2024**

**LEMBAR PENGESAHAN LAPORAN PENELITIAN  
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT  
UNIVERSITAS TERBUKA**

1.	a.	Judul Penelitian	: Inovasi Software Evaluation Pada Aplikasi Mobile Melalui Kombinasi Usability Testing Dan Sentiment Analysis
	b.	Skema Penelitian	: KOM - Dosen Pemula
	c.	Area Penelitian	: Kompetitif Keilmuan
2.		Ketua Peneliti	
	a.	Nama Lengkap & Gelar	: Hasan Basri, S.kom., M.Kom.
	b.	NIP/NIDN	: 199205172024061003/ 0317059202
	c.	Golongan Kepangkatan	: III/B
	d.	Jabatan Akademik	: Dosen Asisten Ahli (Masa Percobaan)
	e.	Fakultas	: FST
	f.	Unit Kerja	: FST
	g.	Program Studi	: S1 Sistem Informasi
3.		Anggota	
	a.	Nama Anggota I	: Irpan Kusyadi, M.Kom.
	b.	NIP Anggota I	: 199010112024061002/ 0411109001
4	a.	Tahun Penelitian	: 2024
	b.	Lama Penelitian	: 3 (tiga) Bulan
5		Biaya Penelitian	
	a.	Diusulkan	: 8.870.000
	b.	Disetujui	: 8.860.000
6		Sumber Biaya	: Universitas Terbuka
7		Pemanfaatan Hasil Penelitian	
	a.	Seminar	: Nasional/Regional/Internasional***)
	b.	Jurnal	: UT/Nasional/Internasional***)
8		Luaran Penelitian	: Jurnal Multitek Indonesia

Tangerang Selatan, 25 November 2024

Mengetahui,  
Dekan FST



Dr. Subekti Nurmawati, M.Si.  
NIP. 196705181991032001

Ketua Peneliti

Hasan Basri, S.Kom., M.Kom.  
NIP. 199205172024061003

Menyetujui,  
Ketua LPPM-UT



Prof. Dewi Artati Padmo Putri, Ph.D.  
NIP. 196107241987102003

Kepala Pusat Penelitian Keilmuan



Dr. Mery Noviyanti, S.Si., M.Pd.  
NIP. 198111242005012003

## DAFTAR ISI

<b>LEMBAR PENGESAHAN LAPORAN PENELITIAN.....</b>	<b>2</b>
<b>DAFTAR ISI .....</b>	<b>3</b>
<b>DAFTAR GAMBAR.....</b>	<b>4</b>
<b>DAFTAR TABEL .....</b>	<b>1</b>
<b>ABSTRAK .....</b>	<b>2</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>3</b>
<b>1.1 Latar Belakang.....</b>	<b>3</b>
<b>1.2 Rumusan Masalah .....</b>	<b>4</b>
<b>1.3 Tujuan Penelitian.....</b>	<b>4</b>
<b>1.4 Batasan Penelitian.....</b>	<b>4</b>
<b>1.5 Manfaat Penelitian.....</b>	<b>5</b>
<b>1.6 Luaran Penelitian .....</b>	<b>5</b>
<b>1.7 State of The Art.....</b>	<b>5</b>
<b>BAB 2 TINJAUAN PUSTAKA .....</b>	<b>7</b>
<b>2.1 Usability Testing .....</b>	<b>7</b>
<b>2.2 Sentiment Analysis.....</b>	<b>8</b>
<b>2.2.1 SVM.....</b>	<b>10</b>
<b>BAB 3 METODE PENELITIAN.....</b>	<b>12</b>
<b>3.1 Tahapan Penelitian .....</b>	<b>12</b>
<b>3.2 Teknik Pengumpulan Data .....</b>	<b>14</b>
<b>3.3 Lokasi Dan Sample Penelitian .....</b>	<b>14</b>
<b>3.4 Roadmap Penelitian.....</b>	<b>14</b>
<b>3.5 Jadwal Penelitian .....</b>	<b>16</b>
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>17</b>
<b>4.1 Pengumpulan data .....</b>	<b>17</b>
<b>4.2 Data Labeling .....</b>	<b>17</b>
<b>4.3 Preprocessing .....</b>	<b>18</b>
<b>4.4 Kernel Selection .....</b>	<b>19</b>
<b>4.5 Tuning Method .....</b>	<b>21</b>
<b>BAB V KESIMPULAN .....</b>	<b>23</b>
<b>DAFTAR PUSTAKA.....</b>	<b>24</b>
<b>Lampiran Identitas Diri Ketua.....</b>	<b>Error! Bookmark not defined.</b>

**DAFTAR GAMBAR**

Gambar 1. Tahapan Usability Testing .....	8
Gambar 2. Tahapan Sentiment Analysis.....	9
Gambar 3. Tahapan Penelitian.....	12
Gambar 4. Roadmap Penelitian .....	15
Gambar 5. Sentiment distribution based on usability testing criteria .....	18
Gambar 6. Sentiment distribution based on usability testing criteria after preprocessing.....	19
Gambar 7. Mean Accuracy VS C with Standard Deviation .....	21

**DAFTAR TABEL**

Tabel 1. Jadwal kegiatan.....	16
Tabel 2. Preprocessing.....	18
Tabel 3. Kernel.....	20
Tabel 4. The resulting performance .....	22

## ABSTRAK

Dalam era digital, aplikasi mobile memainkan peran signifikan dalam kehidupan sehari-hari, dengan lebih dari 2,4 juta aplikasi tersedia di Google Play Store pada tahun 2023. Persaingan ini mendorong pengembang untuk meningkatkan usability aplikasi demi mempertahankan pengguna. Metode usability testing konvensional seringkali terbatas oleh biaya, waktu, dan sumber daya, sehingga mendorong adopsi sentiment analysis sebagai pendekatan alternatif yang efisien. Penelitian ini menggunakan teknik Support Vector Machine (SVM) untuk menganalisis sentimen dan mengukur usability berdasarkan lima kriteria utama. Data yang digunakan mencakup 2.000 ulasan pengguna aplikasi perbankan, dengan proses pelabelan multilabel oleh dua anotator. Eksperimen menunjukkan bahwa kernel linear SVM memberikan kinerja terbaik, dengan akurasi 70,50%, F1 Score 0,8618, dan Hamming Loss 0,0783. Optimasi parameter C menggunakan Grid Search menghasilkan peningkatan akurasi hingga 75,20%, F1 Score 0,8775, dan Hamming Loss 0,0686 pada nilai C optimal sebesar 0,01. Temuan ini menunjukkan potensi sentiment analysis sebagai alat adaptif untuk usability testing, menawarkan solusi berbasis data yang lebih hemat sumber daya dan relevan dalam konteks dunia nyata. Penelitian ini diharapkan memberikan kontribusi dalam pengembangan aplikasi mobile yang lebih responsif terhadap kebutuhan pengguna dan memperluas penerapan machine learning dalam software engineering.

**Kata Kunci :** *Usability Testing, Sentiment Analysis, Machine Learning for Software Engineering, Mobile Application Evaluation.*

## BAB 1 PENDAHULUAN

### 1.1 Latar Belakang

Dalam era digital, aplikasi mobile telah menjadi bagian penting dari kehidupan sehari-hari. Di Google Play Store, terdapat kurang lebih dari 2.4 juta aplikasi yang mencakup berbagai kategori per 2023 (Ceci, 2024). Persaingan ini memaksa pengembang untuk menghasilkan aplikasi yang tidak hanya inovatif, tetapi juga memiliki usability yang tinggi. Usability menjadi faktor utama yang memengaruhi keberhasilan aplikasi, karena aplikasi yang intuitif dan efisien cenderung memiliki tingkat retensi pengguna yang lebih baik (Wang et al., 2022).

Usability Testing biasanya dilakukan melalui metode konvensional seperti usability testing (kuesioner), heuristic evaluation, dan expert reviews (Galavi et al., 2024) (Zulzalil et al., 2023). Meskipun efektif, metode ini memiliki keterbatasan, terutama dalam hal biaya, waktu, dan sumber daya. Metode pengujian konvensional memerlukan partisipasi aktif dari pengguna dalam skenario terkontrol, yang kadang tidak mencerminkan penggunaan aplikasi di dunia nyata.

Dalam konteks ini, sentiment analysis menawarkan solusi yang lebih efisien dan adaptif (Nandwani & Verma, 2021). Sentiment analysis adalah teknik yang digunakan untuk mengekstraksi opini dan emosi dari teks, seperti ulasan pengguna. Di Google Play Store, pengguna sering memberikan ulasan yang berisi pendapat mereka tentang aplikasi. Ulasan ini, jika dianalisis dengan tepat, dapat memberikan wawasan berharga mengenai aspek usability aplikasi.

Sentiment analysis merupakan bagian dari Natural Language Processing (NLP), sebuah bidang yang telah banyak diterapkan dalam berbagai proses Software Engineering. Salah satu contohnya adalah dalam Software Requirement Engineering, di mana NLP digunakan untuk menganalisis user stories. Dengan menggunakan NLP, pengembang dapat secara otomatis mengekstraksi kebutuhan dan harapan pengguna dari teks naratif, sehingga dapat membantu dalam merancang software yang lebih sesuai dengan kebutuhan pengguna (Necula et al., 2024).

Penerapan sentiment analysis pada ulasan pengguna menawarkan beberapa keunggulan. Teknik ini memungkinkan pengumpulan data secara otomatis dalam skala besar, serta memberikan konteks dunia nyata karena ulasan dibuat oleh pengguna tanpa pengaruh skenario terkontrol. Selain itu, sentiment analysis dapat mengidentifikasi pola dan tren dalam persepsi pengguna, membantu pengembang untuk memahami kekuatan dan kelemahan aplikasi mereka secara lebih cepat.

Namun, penerapan sentiment analysis tidak lepas dari tantangan. Salah satu tantangan utamanya adalah memastikan keandalan dan validitas hasil analisis. Ulasan pengguna sering kali menggunakan bahasa informal, dengan ejaan yang tidak konsisten, serta dapat mengandung ambiguitas atau sarkasme, yang menyulitkan proses analisis otomatis (Nath & Dwivedi, 2024). Selain itu, bias dalam ulasan juga menjadi masalah, karena ulasan cenderung lebih banyak ditulis oleh pengguna yang sangat puas atau sangat tidak puas.

Penelitian ini bertujuan untuk mengevaluasi sejauh mana sentiment analysis dapat diandalkan dalam mengukur aspek usability aplikasi mobile, serta mengidentifikasi tantangan dan keterbatasannya. Melalui pendekatan ini, penelitian diharapkan dapat memberikan solusi yang lebih adaptif dan berbasis data dalam Usability Testing.

Dengan mengintegrasikan sentiment analysis, pengembang dapat merespons kebutuhan pengguna lebih cepat dan lebih tepat. Penelitian ini juga memberikan kontribusi dalam memperluas pemahaman tentang penerapan sentiment analysis di luar konteks tradisionalnya, serta membantu pengembang dalam mengoptimalkan aplikasi berdasarkan data ulasan pengguna.

## **1.2 Rumusan Masalah**

Mengacu pada latar belakang yang telah diuraikan, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

- a. Sejauh mana sentiment analysis dapat diandalkan dalam mengidentifikasi dan mengukur aspek-aspek usability pada aplikasi mobile?
- b. Apa saja tantangan dan keterbatasan dalam menerapkan sentiment analysis sebagai metode Usability Testing berbasis ulasan pengguna di Google Playstore?

## **1.3 Tujuan Penelitian**

Tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

- a. Mengembangkan pendekatan baru untuk Usability Testing aplikasi mobile dengan memanfaatkan ulasan pengguna di Google Playstore melalui sentiment analysis.
- b. Menilai keandalan dan validitas sentiment analysis dalam mengidentifikasi dan mengukur aspek usability pada aplikasi mobile.

## **1.4 Batasan Penelitian**

Batasan dalam penelitian ini adalah sebagai berikut:

- a. Penelitian ini dibatasi pada dua aplikasi mobile yang dipilih dari kategori yang sama atau sejenis di Google Playstore. Kedua aplikasi tersebut harus memiliki volume ulasan yang cukup besar untuk memastikan data yang representatif.
- b. Data komentar yang akan dianalisis dibatasi hanya pada komentar yang diposting selama bulan Agustus 2024.
- c. Hanya komentar yang ditulis dalam bahasa Indonesia yang akan dianalisis untuk memudahkan proses sentiment analysis dan klasifikasi. Komentar yang tidak relevan, seperti komentar yang hanya berisi satu kata atau tidak terkait langsung dengan usability aplikasi, akan dikecualikan.

### **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah sebagai berikut:

- a. Penelitian ini memberikan kontribusi dalam mengembangkan metode baru untuk Usability Testing aplikasi mobile.
- b. Penelitian ini berkontribusi dalam mendorong pendekatan pengembangan aplikasi yang lebih berfokus pada pengguna (user-centric), di mana pengalaman dan kepuasan pengguna menjadi prioritas utama.

### **1.6 Luaran Penelitian**

Dalam penelitian ini penulis ingin memiliki target luaran sebagai berikut:

- a. Terbit dalam Jurnal nasional terakreditasi

### **1.7 State of The Art**

Penelitian ini menawarkan solusi modern dan efisien untuk masalah yang sangat relevan dalam pengembangan aplikasi mobile saat ini. Dengan mengintegrasikan sentiment analysis dalam Usability Testing yang merupakan salah satu tahap evaluasi dalam software Engineering, penelitian ini tidak hanya membantu pengembang dalam memahami pengalaman pengguna dengan lebih baik, tetapi juga memungkinkan untuk merespons kebutuhan pasar dengan lebih cepat dan tepat.

Usability testing telah lama digunakan untuk mengidentifikasi masalah dalam desain antarmuka melalui interaksi langsung dengan pengguna dalam skenario terkontrol. Meskipun efektif, metode ini terbatas dalam hal representasi dunia nyata dan memerlukan sumber daya yang signifikan. Di sisi lain, sentiment analysis memanfaatkan Natural Language Processing (NLP) dan Machine Learning

(ML) untuk menganalisis ulasan pengguna secara otomatis, memberikan wawasan berbasis data yang mencerminkan pengalaman pengguna sebenarnya di luar skenario terkontrol.

Kombinasi kedua pendekatan ini menciptakan evaluasi yang lebih holistik. Usability testing memberikan wawasan mendalam dari interaksi langsung, sementara sentiment analysis memungkinkan pengumpulan data dalam skala besar yang mencerminkan kondisi dunia nyata. Dengan mengintegrasikan kedua metode ini, pengembang dapat lebih cepat mengidentifikasi dan memahami kekuatan serta kelemahan aplikasi, memungkinkan respon yang lebih efektif terhadap kebutuhan pengguna. Pendekatan ini juga membuka peluang untuk inovasi lebih lanjut dalam menciptakan software yang lebih adaptif dan berfokus pada pengguna.

## BAB 2 TINJAUAN PUSTAKA

### 2.1 Usability Testing

Usability merupakan sejauh mana sistem, produk, atau layanan dapat digunakan oleh pengguna tertentu untuk mencapai tujuan tertentu dengan efektivitas, efisiensi, dan kepuasan dalam konteks penggunaan tertentu (Hertzum, 2020). Usability testing adalah salah satu metode penting dalam pengembangan produk, terutama dalam memastikan bahwa sistem atau layanan yang dikembangkan benar-benar sesuai dengan kebutuhan pengguna.

Dalam usability testing, terdapat beberapa metode yang sering digunakan, seperti Think-Aloud Protocols, di mana pengguna diminta untuk menjelaskan apa yang dipikirkan saat menggunakan sistem, Question-Asking Protocols, di mana pengguna diberi pertanyaan selama proses penggunaan, Performance Measurements, fokus pada pengukuran waktu atau tingkat keberhasilan tugas yang dilakukan oleh pengguna, Log Analysis, mempelajari data penggunaan dari log sistem, Eye Tracking, melacak pergerakan mata pengguna untuk memahami area yang paling menarik perhatian, Remote Usability Testing, di mana pengujian dilakukan secara jarak jauh, serta Field Usability Testing, dilakukan di lingkungan nyata di mana sistem akan digunakan (Hertzum, 2020) (Agustianto et al., 2022) (Johnson et al., 2022).

Untuk mengevaluasi hasil dari usability testing, terdapat beberapa kriteria utama yang digunakan (Jakob Nielsen, 1993)

a. Learnability (Kemudahan Belajar)

Kemampuan pengguna baru untuk memahami dan mulai menggunakan sistem dengan cepat. Sistem yang memiliki learnability tinggi akan memungkinkan pengguna untuk mempelajari fungsinya dengan mudah dan tanpa banyak hambatan.

b. Efficiency (Efisiensi)

Seberapa cepat pengguna dapat menyelesaikan tugas setelah mereka memahami cara kerja sistem. Efisiensi ini penting karena berhubungan langsung dengan seberapa produktif pengguna bisa saat menggunakan sistem.

c. Memorability (Kemudahan Mengingat)

Seberapa mudah pengguna dapat kembali menggunakan sistem setelah tidak

menggunakannya dalam jangka waktu tertentu. Sistem yang baik harus memungkinkan pengguna untuk mengingat cara penggunaan tanpa harus memulai dari awal.

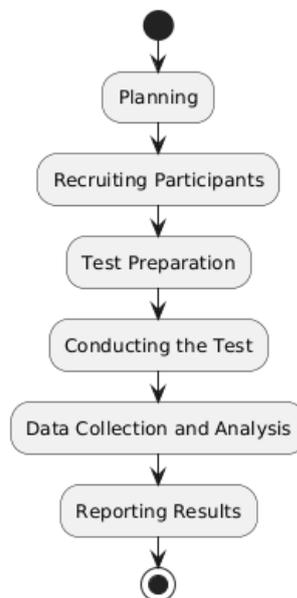
d. Error Tolerance (Toleransi Kesalahan)

Sistem yang toleran terhadap kesalahan akan memberikan petunjuk dan panduan yang jelas agar pengguna dapat kembali ke jalur yang benar.

e. Satisfaction (Kepuasan Pengguna)

Tingkat kepuasan keseluruhan pengguna terhadap sistem, mencakup aspek kenyamanan, kepercayaan, dan perasaan positif lainnya saat menggunakan sistem.

Proses usability testing biasanya dilakukan melalui beberapa tahapan, tahapannya dapat dilihat pada gambar.



Gambar 1. Tahapan Usability Testing

## 2.2 Sentiment Analysis

Sentiment analysis merupakan proses untuk mengidentifikasi dan mengategorikan opini yang terkandung dalam teks, baik itu positif, negatif, atau netral. Sentiment analysis memungkinkan sistem untuk secara otomatis memahami emosi, pandangan, atau perasaan yang diekspresikan oleh pengguna terhadap topik tertentu. Penerapannya mencakup berbagai bidang seperti analisis ulasan produk, pemantauan media sosial, hingga analisis politik, sehingga memberikan wawasan yang sangat berharga dalam pengambilan keputusan (Aftab et al., 2023). Metode-metode sentiment analysis terbagi menjadi tiga pendekatan utama:

a. Lexicon-based

Metode ini menggunakan kamus atau daftar kata yang memiliki nilai sentimen (positif atau negatif). Kata-kata dalam teks kemudian dicocokkan dengan leksikon untuk menentukan polaritas sentimen.

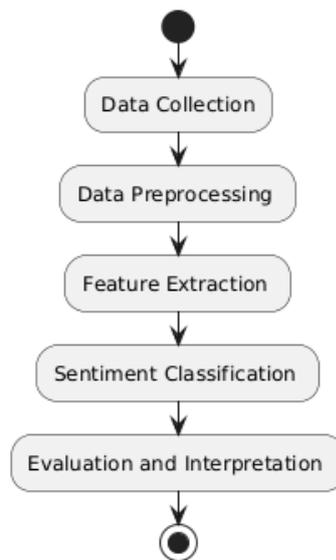
b. Berbasis machine learning

Dalam metode ini, model dilatih menggunakan dataset yang sudah dilabeli dengan kategori sentimen. Algoritma yang sering digunakan antara lain Naive Bayes, Support Vector Machine (SVM), dan Random Forest. Pendekatan ini memanfaatkan fitur-fitur dari teks seperti kata kunci, frekuensi kata, dan n-grams.

c. Hybrid

Menggabungkan metode leksikon dan machine learning untuk mencapai hasil yang lebih akurat dan dapat mengatasi kekurangan dari masing-masing metode.

Tahapan sentiment analysis dilakukan melalui beberapa tahapan, tahapannya dapat di lihat pada gambar:



Gambar 2. Tahapan Sentiment Analysis

a. Data Collection (Pengumpulan Data)

Mengumpulkan data teks dari sumber-sumber yang relevan seperti media sosial, ulasan produk, atau survei.

b. Data Preprocessing (Pra-pemrosesan Data)

Membersihkan data dengan menghilangkan noise seperti stop words, tanda baca, dan normalisasi kata untuk mendapatkan teks yang lebih bersih.

c. Feature Extraction (Ekstraksi Fitur)

Mengambil fitur-fitur penting dari teks seperti kata kunci, frasa, atau n-grams

untuk digunakan dalam proses klasifikasi.

d. Sentiment Classification (Klasifikasi Sentimen)

Menggunakan algoritma machine learning atau pendekatan leksikon untuk mengkategorikan teks menjadi sentimen positif, negatif, atau netral.

e. Evaluation and Interpretation (Evaluasi dan Interpretasi)

Mengukur kinerja model dengan metrik seperti akurasi, precision, recall, dan F1-score, serta menginterpretasikan hasil analisis untuk diaplikasikan dalam pengambilan keputusan.

### 2.2.1 SVM

Support Vector Machine (SVM) adalah salah satu algoritma supervised learning yang banyak digunakan dalam klasifikasi dan regresi. Algoritma ini sangat efektif terutama untuk masalah klasifikasi biner. SVM bekerja dengan mencari hyperplane yang paling baik memisahkan data ke dalam dua kategori. Hyperplane ini adalah garis batas (boundary) yang memisahkan dua kelas dengan margin maksimum, yaitu jarak terbesar antara data dari kedua kelas dengan hyperplane. Semakin besar margin, semakin baik model dalam melakukan klasifikasi (Nugroho & Norhikmah, 2024)

Secara konsep, SVM mencoba untuk menemukan support vectors, yaitu titik-titik data yang paling dekat dengan hyperplane. Titik-titik ini memainkan peran penting karena posisi mereka menentukan posisi hyperplane yang optimal. Ketika data tidak dapat dipisahkan secara linear, SVM menggunakan teknik kernel trick untuk memetakan data ke dimensi yang lebih tinggi, memungkinkan pemisahan yang lebih baik.

Rumus SVM secara matematis untuk sebuah hyperplane dalam dimensi  $n$  dinyatakan sebagai berikut:

$$w \cdot x + b = 0$$

Di mana:

$w$  = vector bobot (weight vector).

$x$  = vector input data.

$b$  = bias atau konstanta.

Tujuan utama SVM adalah memaksimalkan margin antara dua kelas, yang dapat dihitung sebagai:

$$margin = \frac{2}{\|w\|}$$

Untuk memaksimalkan margin, SVM menyelesaikan masalah optimasi berikut:

$$\min \frac{1}{2} \|w\|^2$$

Dengan syarat:

$$y_i(w \cdot x_i + b) \geq 1 \quad \forall i$$

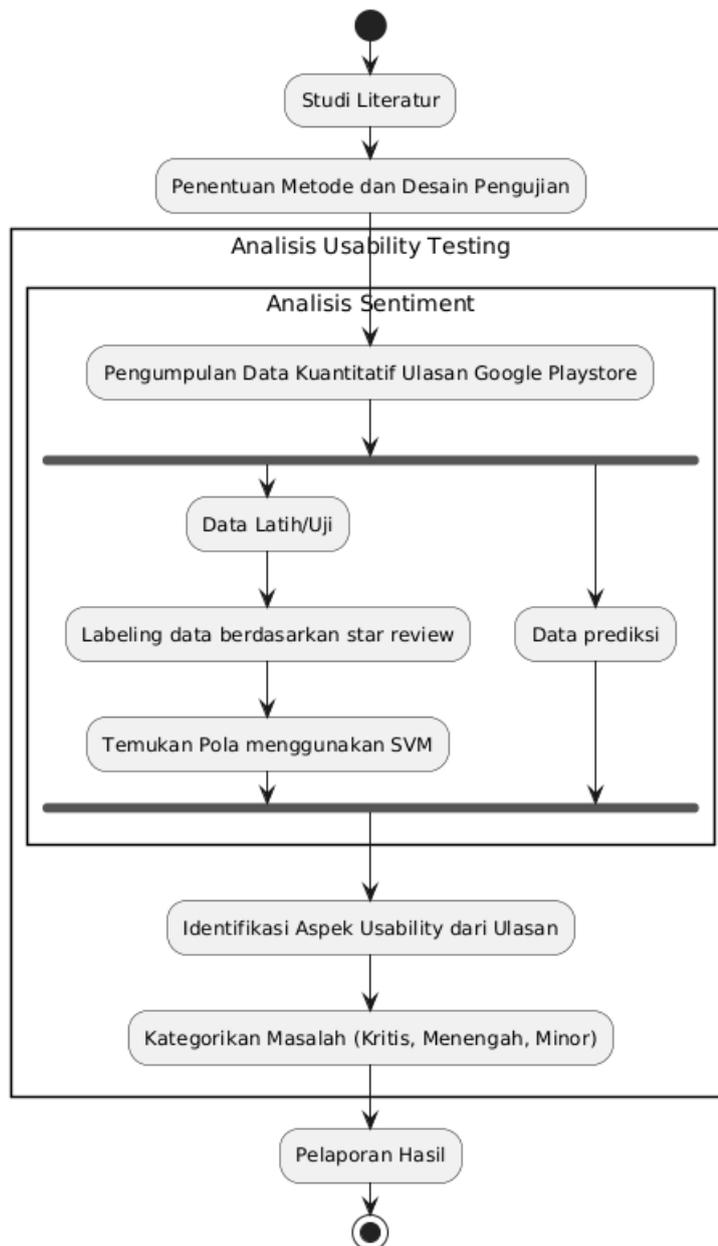
Di mana  $y_i$  adalah label kelas (+1 atau -1) untuk setiap sampel  $x_i$ .

Jika data tidak terpisahkan secara linear, fungsi kernel seperti linear, polynomial, radial basis function (RBF), atau sigmoid digunakan untuk mengubah ruang data sehingga dapat dipisahkan dengan hyperplane

## BAB 3 METODE PENELITIAN

### 3.1 Tahapan Penelitian

Tahapan penelitian dalam usability testing menggunakan sentiment analysis ini dapat dilihat pada gambar 3.1 sebagai berikut:



Gambar 3. Tahapan Penelitian

Tahapan penelitian yang telah digambarkan pada gambar 3.1 dapat jelaskan sebagai berikut:

#### 1. Studi Literatur

Tahap ini melibatkan penelitian dan kajian literatur untuk memahami teori, metode, dan pendekatan yang relevan dengan usability testing dan sentiment

analysis. Penelitian ini bertujuan untuk memperoleh pemahaman yang mendalam tentang konsep-konsep kunci dan bagaimana metode yang ada dapat diadaptasi.

## 2. Penentuan Metode dan Desain Pengujian

Setelah studi literatur, langkah berikutnya adalah menentukan metode dan desain pengujian yang akan digunakan dalam penelitian. Ini mencakup pemilihan teknik sentiment analysis, algoritma klasifikasi (misalnya SVM), dan cara pengumpulan serta pengolahan data.

## 3. Analisis Usability Testing

Tahap ini merupakan proses inti dari penelitian, yang terdiri dari beberapa sub-proses:

### a. Analisis Sentiment

Pengumpulan Data Kuantitatif Ulasan Google Playstore. Pada tahap ini, data ulasan dari Google Playstore dikumpulkan sebagai basis data untuk analisis. Data ini mencakup ulasan pengguna yang mencerminkan pengalaman pengguna dengan aplikasi.

### b. Data Latih/Uji:

Data ulasan dipisahkan menjadi data latih dan data uji. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengukur akurasi model.

### c. Labeling Data Berdasarkan Star Review:

Data ulasan diberi label berdasarkan jumlah bintang (star review) yang diberikan oleh pengguna. Label ini digunakan sebagai acuan dalam training model. Selain star review, Pelabelan akan ditinjau ulang secara manual guna pengecekan sehingga mendapatkan hasil akurasi yang maksimal

### d. Membangun Pola Menggunakan SVM:

Support Vector Machine (SVM) digunakan untuk menemukan pola dalam data ulasan yang terkait dengan usability. Pola ini kemudian digunakan untuk memprediksi sentimen ulasan baru.

### e. Data Prediksi:

Data ulasan baru yang belum diberi label dianalisis menggunakan model yang sudah dilatih untuk memprediksi sentimen dan mengidentifikasi aspek usability yang relevan.

#### 4. Identifikasi Aspek Usability dari Ulasan

Setelah analisis sentiment dilakukan, tahap ini melibatkan identifikasi aspek-aspek spesifik dari usability yang disebutkan dalam ulasan pengguna. Ini dapat mencakup elemen-elemen seperti kemudahan penggunaan, kecepatan aplikasi, dan desain antarmuka.

#### 5. Kategorikan Masalah (Kritis, Menengah, Minor)

Setelah mengidentifikasi aspek usability, masalah-masalah yang ditemukan dikategorikan berdasarkan tingkat keparahannya: kritis, menengah, atau minor. Kategorisasi ini membantu dalam prioritas perbaikan dan peningkatan aplikasi.

#### 6. Pelaporan Hasil

Tahap akhir dari proses ini adalah pelaporan hasil penelitian. Laporan ini mencakup temuan utama dari analisis sentiment dan usability, serta rekomendasi perbaikan metode berdasarkan masalah yang teridentifikasi

### **3.2 Teknik Pengumpulan Data**

Data yang digunakan dalam penelitian ini adalah ulasan pengguna dari Google Playstore. Teknik pengumpulan data ini dilakukan dengan menarik data ulasan secara langsung dan mencatatnya secara sistematis untuk dianalisis. Data yang diambil meliputi nama pengguna, waktu ulasan diberikan, isi ulasan, dan rating bintang yang diberikan oleh pengguna. Ulasan-ulasan ini kemudian akan dianalisis menggunakan teknik sentiment analysis untuk mengevaluasi usability aplikasi mobile. Data ulasan yang terkumpul akan menjadi sumber utama untuk menilai pengalaman dan kepuasan pengguna terhadap aplikasi yang diteliti.

### **3.3 Lokasi Dan Sample Penelitian**

Penelitian ini akan dilaksanakan di Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Terbuka dengan sampel ulasan pengguna aplikasi mobile pada google playstore.

### **3.4 Roadmap Penelitian**

Penelitian ini dilakukan selama 3 tahun sehingga *roadmap* penelitian dapat dilihat sebagai berikut:



Gambar 4. Roadmap Penelitian

Roadmap penelitian ini bertujuan untuk mengembangkan dan mengoptimalkan metode usability testing pada aplikasi mobile dengan memanfaatkan teknologi canggih seperti Sentiment Analysis, SVM, RNN, dan data interaksi pengguna. Pada tahun pertama, fokusnya adalah membangun dasar pemahaman tentang penggunaan Sentiment Analysis dan SVM untuk mengukur dan memprediksi sentimen dari ulasan pengguna. Tahun kedua akan melanjutkan dengan mengoptimalkan pengujian usability menggunakan kombinasi Sentiment Analysis dan RNN untuk meningkatkan akurasi klasifikasi sentimen. Pada tahun ketiga, penelitian akan memperluas pendekatan dengan mengintegrasikan data multimodal, yaitu ulasan teks dan data interaksi pengguna, untuk menciptakan model pengujian usability yang lebih komprehensif dan efektif. Dengan pendekatan ini, diharapkan model yang dihasilkan mampu mendeteksi masalah usability lebih cepat dan memberikan wawasan yang lebih mendalam tentang kebutuhan dan kepuasan pengguna, sehingga meningkatkan kualitas aplikasi mobile secara keseluruhan.



## **BAB IV HASIL DAN PEMBAHASAN**

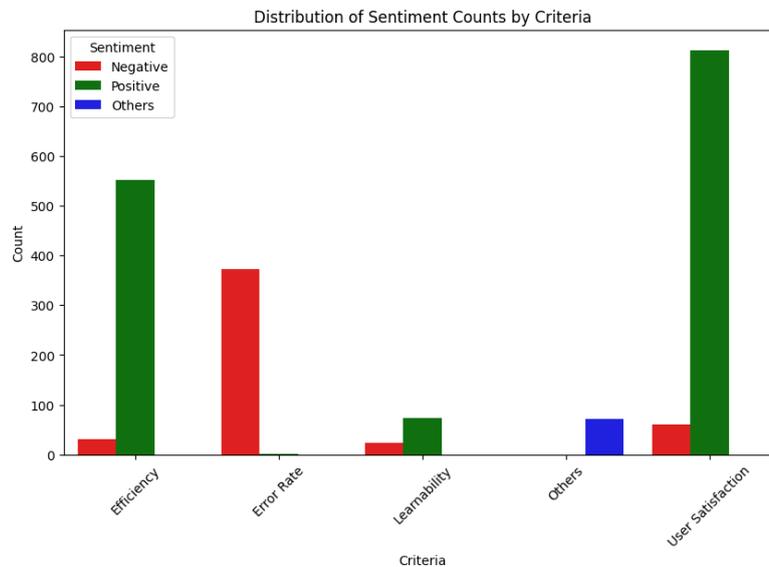
### **4.1 Pengumpulan data**

Penelitian ini menggunakan data ulasan dari salah satu aplikasi perusahaan perbankan yang tersedia di Google Playstore, dengan total sekitar 1,5 juta ulasan. Jumlah ulasan yang sangat besar ini memberikan potensi yang signifikan untuk digunakan dalam penelitian yang berkaitan dengan analisis sentimen dan pengujian kegunaan. Data tersebut mencerminkan pendapat pengguna mengenai pengalaman mereka dalam menggunakan aplikasi perbankan, baik dari segi kepuasan, efisiensi, maupun kesalahan yang mungkin terjadi.

Pada tanggal 24 Agustus 2024, sebanyak 2.000 ulasan diambil dari total dataset untuk dianalisis lebih lanjut berdasarkan kriteria usability testing. Proses pelabelan dilakukan dengan mengikuti 5 kriteria usability testing (Weichbroth, 2020) (Shareef & Khan, 2019). Lima aspek utama yang digunakan dalam penelitian ini adalah Efisiensi, Kepuasan Pengguna, Learnability, Memorability, dan Tingkat Kesalahan.

### **4.2 Data Labeling**

Penelitian ini menggunakan pendekatan multi-label dalam melabeli data ulasan, dengan dua orang anotator yang berperan dalam proses ini. Annotator 'A' bertanggung jawab untuk melabeli ulasan berdasarkan analisis sentimen dengan dua kategori, yaitu Positif dan Negatif. Setelah label sentimen ditetapkan, Annotator 'B' melabeli ulasan berdasarkan kriteria pengujian kegunaan, yang mencakup lima aspek utama: Kemudahan dipelajari, Efisiensi, Daya Ingat, Tingkat Kesalahan, dan Kepuasan Pengguna. Selain itu, Annotator 'B' juga melakukan pengecekan terhadap label sentimen yang telah diberikan oleh Annotator 'A' untuk memastikan tidak ada kesalahan pelabelan. Tahap terakhir dalam proses pelabelan ini adalah Annotator 'A' kembali melakukan pengecekan terhadap label yang telah diberikan untuk Usability Testing Criteria oleh Annotator 'B' untuk memastikan konsistensi dan ketepatannya.



Gambar 5. *Sentiment distribution based on usability testing criteria*

Berdasarkan data yang diperoleh dari hasil pelabelan ulasan, dapat disimpulkan bahwa distribusi ulasan pada kriteria usability testing dan analisis sentimen adalah sebagian besar ulasan yang mengomentari kriteria efficiency juga didominasi oleh ulasan positif (552 ulasan). Sedangkan kriteria Error Rate yang berkaitan dengan kesalahan dalam menggunakan aplikasi memiliki sentimen negatif (373 ulasan), yang mengindikasikan bahwa pengguna banyak mengalami masalah teknis atau kesalahan. Kriteria Learnability menunjukkan bahwa aplikasi ini cukup mudah untuk dipelajari, dengan 74 ulasan positif. Selanjutnya, Kepuasan Pengguna, kriteria ini mendominasi dengan 813 ulasan positif, yang mencerminkan tingkat kepuasan yang tinggi terhadap aplikasi.

Namun, pada hasil pelabelan, kriteria Memorability tidak ditemukan. Hal ini dapat mengindikasikan bahwa pengguna tidak secara eksplisit mengomentari kemudahan mengingat cara menggunakan aplikasi, atau kriteria ini tidak relevan dalam konteks ulasan. Hal ini dapat menjadi temuan penting dalam memahami fokus utama pengguna terhadap kegunaan aplikasi. Sebanyak 72 ulasan lainnya memberikan ulasan di luar konteks aplikasi yang digunakan.

### 4.3 Preprocessing

Setelah menganalisis distribusi data, beberapa kombinasi kriteria dan sentimen dengan jumlah ulasan yang sedikit akan dieliminasi untuk menjaga relevansi data dalam penelitian. Kombinasi yang dieliminasi adalah sebagai berikut:

Tabel 2. *Preprocessing*

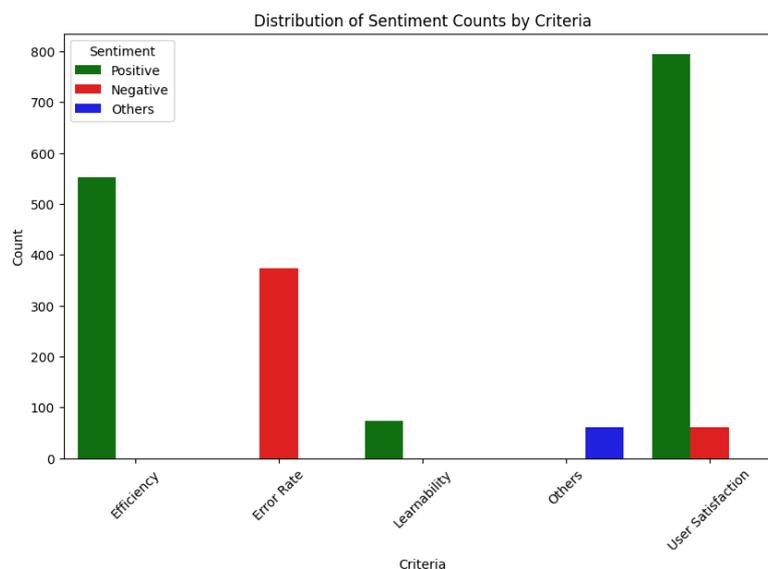
Combination	
Error Rate + Positive	(1 review)

Efficiency + Negative	(30 reviews)
Learnability + Negative	(24 reviews)

Kombinasi ini dihapus karena jumlah data yang sedikit, yang dapat mempengaruhi kualitas analisis dan generalisasi model. Dengan demikian, data yang tersisa mencakup distribusi ulasan yang lebih signifikan untuk mendukung analisis kegunaan dan analisis sentimen secara lebih efektif.

Pada tahap preprocessing, serangkaian langkah dilakukan untuk membersihkan dan mempersiapkan data ulasan pengguna agar konsisten dan siap untuk dianalisis. Proses ini dimulai dengan mengubah angka menjadi kata-kata, misalnya angka '1' diubah menjadi 'satu'. Selanjutnya, karakter non-ASCII yang tidak relevan dihilangkan untuk memastikan data lebih terorganisir dan mudah diproses.

Selain itu, dilakukan juga lemmatization, yaitu mengubah setiap kata menjadi bentuk dasarnya, misalnya kata 'membeli' diubah menjadi 'beli'. Lemmatization penting dilakukan untuk mengurangi variasi kata yang sebenarnya memiliki arti yang sama, sehingga model dapat lebih mudah mengenali pola dalam data teks. Langkah terakhir adalah pembersihan tambahan, yang meliputi penghapusan tanda baca, spasi yang berlebihan, dan karakter-karakter yang tidak relevan. Proses ini bertujuan untuk memastikan bahwa data yang digunakan dalam analisis benar-benar bersih dan siap untuk diproses lebih lanjut.



Gambar 6. *Sentiment distribution based on usability testing criteria after preprocessing*

#### 4.4 Kernel Selection

Langkah selanjutnya dalam penelitian ini adalah pemilihan kernel untuk

model Support Vector Machine (SVM). Pemilihan kernel sangat penting karena kernel menentukan bagaimana data dipetakan ke ruang dimensi yang lebih tinggi, sehingga mempengaruhi kinerja model dalam tugas klasifikasi. Pemilihan kernel dilakukan berdasarkan hasil percobaan awal dengan parameter default untuk mengevaluasi kinerja dasar dari setiap kernel.

Tabel 3. *Kernel*

Kernel	Accuracy	Hamming Loss	F1 Score	Precision	Recall
Linear	70.50 %	0.0783	0.8618	0.8573	0.8664
RBF	64.23 %	0.1037	0.8075	0.8474	0.7712
Sigmoid	68.41 %	0.0783	0.8564	0.8867	0.8200
Poly	42.30 %	0.2167	0.6145	0.6165	0.6124

Kernel linear memberikan hasil terbaik secara keseluruhan, dengan nilai akurasi tertinggi (70.50%), F1 Score tertinggi (0.8618), dan Hamming Loss terendah (0.0783). Hal ini menunjukkan bahwa kernel linear paling efektif dalam mengklasifikasikan data ulasan, terutama dalam memaksimalkan keseimbangan antara precision dan recall. Model ini berhasil meminimalkan kesalahan dan memberikan margin pemisahan yang optimal. Kernel RBF juga berkinerja cukup baik, terutama dalam hal presisi (0.8474). Namun, recall yang lebih rendah (0.7712) menunjukkan bahwa kernel ini mengalami kesulitan dalam mengidentifikasi ulasan secara menyeluruh, sehingga menghasilkan lebih banyak kesalahan daripada kernel linear. Kernel Sigmoid memiliki presisi tertinggi (0.8867), yang menunjukkan kemampuan yang baik untuk memprediksi label positif dengan benar. Namun, recall yang lebih rendah dibandingkan dengan kernel linear (0.8200) menunjukkan bahwa model ini kurang komprehensif dalam menangkap semua ulasan dengan benar, seperti yang terlihat dari akurasi yang lebih rendah (68.41%). Sementara itu, Kernel Polynomial menunjukkan kinerja terendah pada semua metrik yang diuji. Dengan akurasi terendah (42.30%), F1 Score yang rendah (0.6145), dan Hamming Loss tertinggi (0.2167), kernel ini kurang cocok untuk data yang digunakan dalam penelitian ini. Model ini mengalami kesulitan dalam mengklasifikasikan ulasan secara akurat dan konsisten.

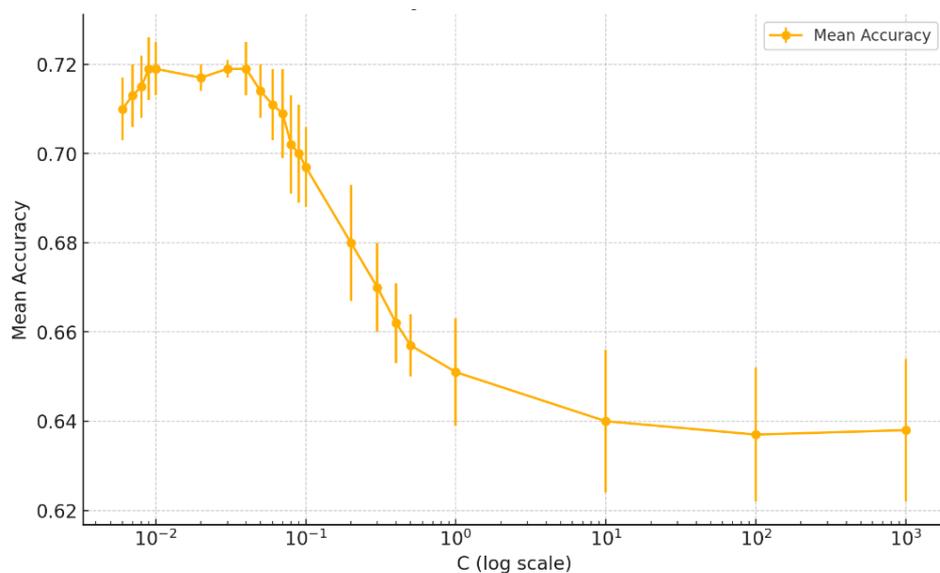
Berdasarkan hasil perbandingan, Linear Kernel adalah yang terbaik untuk data yang digunakan dalam penelitian ini. Kernel ini memberikan akurasi tertinggi, F1 Score tertinggi, dan Hamming Loss terendah, yang mengindikasikan bahwa model ini mampu memisahkan data secara optimal dan menghasilkan performa terbaik dibandingkan dengan kernel lainnya. Kernel RBF dan Sigmoid juga menunjukkan performa yang kompetitif, terutama dalam hal presisi, tetapi masih kalah dalam hal recall dan akurasi secara keseluruhan.

#### 4.5 Tuning Method

Pada penelitian ini, metode tuning menggunakan Grid Search dilakukan untuk mencari nilai  $C$  yang optimal pada model SVM dengan kernel linear. Berdasarkan hasil pemilihan kernel yang dilakukan sebelumnya, kernel linear dipilih karena memberikan performa terbaik dalam mengklasifikasikan data ulasan berbasis teks dengan representasi TF-IDF. Kernel ini telah terbukti mampu menangani data berdimensi tinggi secara efektif, sehingga proses tuning hanya difokuskan pada pengoptimalan hyperparameter  $C$ .

Nilai  $C$  adalah parameter penting dalam SVM, yang mengontrol trade-off antara margin yang besar dan klasifikasi yang tepat pada data training. Dalam Grid Search, beberapa nilai  $C$  diuji untuk menemukan kombinasi yang memberikan akurasi terbaik. Proses ini dilakukan dengan validasi silang untuk memastikan bahwa model tidak overfitting dan dapat memberikan generalisasi yang baik pada data uji.

Proses tuning parameter dilakukan dengan Grid Search untuk mencari nilai  $C$  yang paling optimal pada model SVM dengan kernel linear. Pada tahap awal, nilai  $C$  yang diujicobakan adalah 0.1, 1, 10, 100, dan 1000, dengan menggunakan 3-fold cross-validation untuk memastikan hasil yang didapatkan tidak overfitting. Dari hasil tuning tersebut, didapatkan bahwa parameter terbaik adalah  $C = 0.1$  yang memberikan performa terbaik pada data yang digunakan.



Gambar 7. Mean Accuracy VS  $C$  with Standard Deviation

Selanjutnya, proses tuning diperluas dengan menguji nilai  $C$  antara 0.1

hingga 0.5 ([0.1, 0.2, 0.3, 0.4, 0.5]), namun tetap menunjukkan bahwa  $C = 0.1$  adalah yang paling optimal. Untuk memeriksa lebih lanjut keakuratan parameter optimal, Grid Search dilakukan lagi dengan rentang  $C$  yang lebih kecil ([0.05, 0.06, 0.07, 0.08, 0.09, 0.1]), dan nilai optimal berubah menjadi  $C = 0.05$ . Pada tahap akhir, tuning disempurnakan dengan menguji nilai yang lebih rendah, yaitu antara 0,01 hingga 0,05 ([0,01, 0,02, 0,03, 0,04, 0,05]). Dari hasil tersebut, didapatkan bahwa nilai  $C$  yang paling optimal adalah 0.01 karena ketika dilakukan percobaan lain dengan nilai  $C$  yang lebih kecil ([0.006, 0.007, 0.008, 0.009, 0.01]), nilai  $C$  yang optimal tetap 0.01. Evaluasi model dengan nilai  $C$  ini menghasilkan kinerja sebagai berikut:

Tabel 4. *The resulting performance*

No	Performance	Value
1	Accuracy SVM (multi-label):	75.20%
2	Hamming Loss (which shows a minimal number of misclassifications)	0.0686
3	F1 Score (shows a good balance between precision and recall)	0.8775
4	Precision	0.8834
5	Recall	0.8717

Hasil ini menunjukkan bahwa  $C = 0,01$  adalah nilai yang paling optimal untuk tugas klasifikasi multi-label dalam penelitian ini. Nilai  $C$  yang kecil ini memungkinkan margin yang lebih besar pada model, yang membantu mencegah overfitting dan meningkatkan generalisasi model ke data uji. Dengan parameter ini, model SVM berhasil bekerja dengan sangat baik dalam klasifikasi sentimen dan pengujian kriteria kegunaan.

## BAB V KESIMPULAN

Penelitian ini menggunakan Support Vector Machine (SVM) untuk melakukan analisis sentimen dan pengujian kegunaan pada ulasan aplikasi mobile dari Google Play Store. Setelah melalui proses pengumpulan data, pelabelan, preprocessing, pemilihan kernel, dan tuning parameter. Data yang diambil sebanyak 2.000 ulasan dari total 1,5 juta ulasan yang mencerminkan berbagai pengalaman pengguna, dengan fokus pada 6 aspek kriteria usability testing. Namun, kriteria Memorability tidak ditemukan dalam ulasan yang dianalisis, yang mengindikasikan bahwa pengguna tidak memberikan komentar secara eksplisit terkait daya ingat aplikasi. Proses pelabelan dilakukan secara multi-label oleh dua anotator, dengan pelabelan anotator 'A' berdasarkan analisis sentimen (Positif dan Negatif), dan anotator 'B' berdasarkan kriteria pengujian kegunaan. Hal ini memungkinkan ulasan dianalisis secara komprehensif dari dua sudut pandang, yaitu aspek sentimen dan kegunaan.

Setelah melakukan percobaan dengan beberapa kernel (Linear, RBF, Sigmoid, Polynomial), hasilnya menunjukkan bahwa Kernel Linear memberikan performa terbaik dengan akurasi 70.50%, F1 Score 0.8618, dan Hamming Loss 0.0783. Kernel linear terbukti paling efektif dalam mengklasifikasikan data ulasan, terutama dalam memaksimalkan keseimbangan antara precision dan recall. Melalui Grid Search untuk mencari nilai C yang optimal, didapatkan bahwa nilai terbaik adalah  $C = 0.01$ , menghasilkan akurasi 75.20%, F1 Score 0.8775, dan Hamming Loss 0.0686 pada data uji.

Percobaan dengan nilai C yang lebih besar dari 0.01 menunjukkan bahwa rata-rata akurasi menurun seiring dengan meningkatnya nilai C, seperti yang terlihat pada hasil rata-rata akurasi sebesar 71.4% pada  $C = 0.05$  dan terus menurun hingga 69.7% pada  $C = 0.1$ . Hal ini menunjukkan bahwa nilai C yang lebih besar membuat model terlalu fokus pada data pelatihan, yang berpotensi menyebabkan overfitting yang mengurangi performa pada data uji. Sebaliknya, percobaan dengan nilai C yang lebih kecil dari 0,01 juga menghasilkan penurunan akurasi. Nilai C yang terlalu kecil menyebabkan model kehilangan kemampuan untuk menangkap pola yang relevan, sehingga mengurangi kemampuan klasifikasi. Dengan demikian, nilai C yang optimal sebesar 0.01 memberikan keseimbangan terbaik antara akurasi dan generalisasi model, tanpa terlalu membatasi atau memperluas margin pemisahan dalam klasifikasi.

## DAFTAR PUSTAKA

- Aftab, F., Bazai, S. U., Marjan, S., Baloch, L., Aslam, S., Amphawan, A., & Neo, T. K. (2023). A Comprehensive Survey on Sentiment Analysis Techniques. *International Journal of Technology*, 14(6), 1288. <https://doi.org/10.14716/ijtech.v14i6.6632>
- Agustianto, K., Utomo, A. H., Ayuninghemi, R., Destarianto, P., Wiryawan, I. G., & Mulyadi, E. (2022). *Eye Tracking Usability Testing Using User-Centered Design Analysis Method*. <https://doi.org/10.2991/assehr.k.220207.045>
- Ceci, L. (2024). *Number of available applications in the Google Play Store from March 2017 to June 2024*. Statista. <https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/>
- Galavi, Z., Norouzi, S., & Khajouei, R. (2024). Heuristics used for evaluating the usability of mobile health applications: A systematic literature review. *DIGITAL HEALTH*, 10. <https://doi.org/10.1177/20552076241253539>
- Hertzum, M. (2020). Usability Testing: A Practitioner's Guide to Evaluating the User Experience. In *Synthesis Lectures on Human-Centered Informatics* (Vol. 1, Issue 1). <https://doi.org/10.2200/S00987ED1V01Y202001HCI045>
- Jakob Nielsen. (1993). *Usability Engineering*. Elsevier. <https://doi.org/10.1016/C2009-0-21512-1>
- Johnson, S. G., Potrebny, T., Larun, L., Ciliska, D., & Olsen, N. R. (2022). Usability Methods and Attributes Reported in Usability Studies of Mobile Apps for Health Care Education: Scoping Review. *JMIR Medical Education*, 8(2), e38259. <https://doi.org/10.2196/38259>
- Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 81. <https://doi.org/10.1007/s13278-021-00776-6>
- Nath, D., & Dwivedi, S. K. (2024). Aspect-based sentiment analysis: approaches, applications, challenges and trends. *Knowledge and Information Systems*. <https://doi.org/10.1007/s10115-024-02200-9>
- Necula, S.-C., Dumitriu, F., & Greavu-Şerban, V. (2024). A Systematic Literature Review on Using Natural Language Processing in Software Requirements Engineering. *Electronics*, 13(11), 2055. <https://doi.org/10.3390/electronics13112055>
- Nugroho, A. W., & Norhikmah, N. (2024). Sentiment Analysis using the Support Vector Machine Algorithm on Covid\_19. *SISTEMASI*, 13(4), 1758. <https://doi.org/10.32520/stmsi.v13i4.3778>
- Shareef, S., & Khan, M. N. A. (2019). Evaluation of Usability Dimensions of Smartphone Applications. *International Journal of Advanced Computer Science and Applications*, 10(9). <https://doi.org/10.14569/IJACSA.2019.0100956>
- Wang, Y., Wang, J., Zhang, H., Ming, X., Shi, L., & Wang, Q. (2022). Where is your app frustrating users? *Proceedings of the 44th International Conference on Software Engineering*, 2427–2439. <https://doi.org/10.1145/3510003.3510189>
- Weichbroth, P. (2020). Usability of Mobile Applications: A Systematic Literature Study. *IEEE Access*, 8, 55563–55577. <https://doi.org/10.1109/ACCESS.2020.2981892>
- Zulzalil, H., Rahmat, H., Ghani, A. A. A., & Kamaruddin, A. (2023). Expert Review on Usefulness of an Integrated Checklist-based Mobile Usability Evaluation Framework. *Journal of Computer Science Research*, 5(3), 57–73. <https://doi.org/10.30564/jcsr.v5i3.5816>

